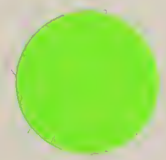
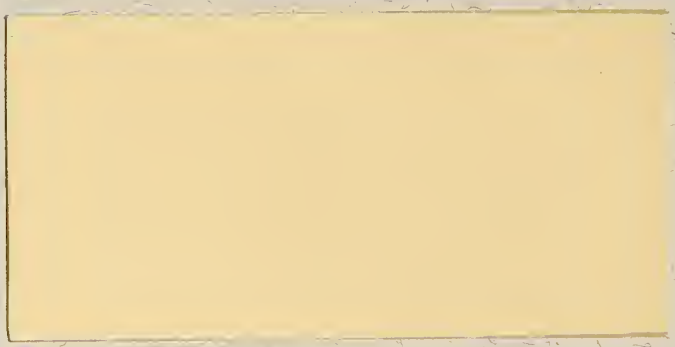
WORKING PAPER: 1403-04 February 1984

Medicaid Claims Data as a Source of Hospital Diagnostic Information

Carol Hamcke



THE URBAN INSTITUTE
WASHINGTON, D.C.



1403-04 February 1984

Medicaid Claims Data as a Source of Hospital Diagnostic Information

Carol Hamcke

This research was supported by Grant No. 18-P-97516/3-01, Health Care Financing Administration, U.S. Department of Health and Human Services. Opinions expressed are those of the author and do not necessarily represent the view of The Urban Institute or its sponsors.

THE URBAN INSTITUTE
2100 M Street, N.W., Washington, D. C. 20037

TABLE OF CONTENTS

	Page
MEDICAID CLAIMS AS A DATA SOURCE	1
DATA RELIABILITY	2
DESCRIPTIVE STATISTICS ON THE CALIFORNIA MEDICAID SAMPLES	4
CONCLUSION	6

LIST OF TABLES

Table		Page
1	Comparison of Diagnostic Data of a California Female AFDC Medicaid Sample and a Nationwide Female Sample 1976	5
2	Hospital Data for a Sample of Medicaid Eligibles by Eligibility Status, Type of Service, and Predominant Source of Care, California 1976-1978	8
3	Distribution of Fifteen Most Frequent Hospital Diagnoses by Predominant Source of Care for Female AFDC Medicaid Eligibles, California 1976-1978	9
4	Categorization of Physician Discretion Based on Hospital Discharge Diagnosis	10
5	Distribution of Hospital Discharge Diagnosis Groupings for California AFDC and Disabled Medicaid Eligibles, 1976-1978	11
6	Distribution of Hospital Diagnosis Groupings by Predominant Source of Care, California Female AFDC Medicaid Recipients, California 1976-1978	12
7	Distribution of Hospital Diagnoses Groupings by Predominant Source of Care, California Disabled Medicaid Recipients, California, 1976-1978	13

The final report of which this paper is a part presents the results of several analyses of medical care usage patterns using a data base drawn from Medicaid claims data for California, classified by diagnosis. It is, thus, appropriate to discuss the advantages and disadvantages of Medicaid claims information as a data source, assess its generalizability by comparing it with diagnostic information from the National Health Survey, and present some descriptive statistics on the medical care usage patterns of the Medicaid eligibles in the study sample. The three sections of this paper address these three issues, respectively.

MEDICAID CLAIMS AS A DATA SOURCE

The obvious advantage of using Medicaid claims records as a data source is that they provide information on large samples of eligibles at a fraction of the cost of collecting comparable data by other methods. For the California Medicaid (Medi-Cal) study, for example, the analysis files contained a record of the medical utilization history of 50,000 eligibles for the first six-month period of each of three study years, 1976-1978, yielding a total sample of 150,000. For much of the analysis a subsample of 7,000 beneficiaries was used, yielding about 2,100 hospitalizations. The histories included expenditures and frequency of service use, hospital discharge diagnoses for the first three hospitalizations in each period, and lengths of stay.

The major disadvantage of Medicaid claims data is that diagnoses are the only means of identifying the reason for health care use and, as such, do not include physician rankings of urgency or severity. This is a substantial disadvantage, because the same discharge diagnosis can cover a wide range of severity which can be an important factor influencing type and cost of

treatment. This disadvantage is counterbalanced to some degree, however, because hospital discharge diagnoses (as classified in the International Classification of Diseases, adapted, eighth edition) provide a less subjective assessment of illness than the more refined methods that use episode-specific information as assessed by a particular physician.

DATA RELIABILITY

There are two facets to the issue of data reliability, both of which will be dealt with briefly in turn. The first is whether the data base used in the analysis provides an accurate reflection of the actual medical care usage characteristics of the sample. The second is whether the sample itself is representative of other samples of policy interest.

There are two respects in which the way the data were compiled might have imparted bias to the results. The first is that, although the analysis file contains information on the discharge diagnosis of the first three inpatient claims in a six-month period, the diagnosis classification methodology used for the study uses only the first such claim. To assess how much of a limitation this imposes on the generalizability of the analysis we calculated, for the AFDC female subsample, the percent of eligibles hospitalized by number of hospitalizations. In a sample of 22,000 eligibles, over 90 percent were not hospitalized at all; 7.59 percent were hospitalized once; only 1.35 percent were hospitalized twice, and less than half of one percent more than twice. Since more than 8 out of 10 of those hospitalized were hospitalized only once, first hospitalizations should yield a relatively representative distribution of diagnoses leading to hospital admission for the sample as a whole.

The second possibly limiting characteristic of the data based used for the analysis has to do with assigning a predominant source of (ambulatory)

care (PSC) for each sample member. The issue is whether our assignment methodology led to an inflated estimate of the proportion of the sample whose PSC was the outpatient department of a hospital. An example will clarify the potential danger.

Let us suppose a beneficiary has had two visits to a private physician's office and one visit to an OPD during the first four months of the study period. The beneficiary is hospitalized during the fifth month. Follow-up visits are required to check the beneficiary's progress, but the physician wants to see the patient in the OPD (because of easy access to equipment, no incurring of overhead expenses, or merely convenience). If two such visits are required, then at the end of the six-month study period the beneficiary will have a total of three physician visits in an OPD and two physician visits in a private office. The PSC assignment methodology we used would, in such a case, wrongly assign the OPD as this sample member's PSC. To assess the extent of this danger, we looked at the number of cases in which, after hospitalization, a patient visited the relevant OPD. The general pattern of OPD visits over time shows that this was not a frequent occurrence.

To assess the representativeness of the diagnostic data on the California Medicaid claims records as a guide to the diagnostic characteristics of other populations, we compared the rank ordering of hospital discharge diagnoses for our AFDC female sample with that of the female sample members of the National Health Survey who were hospitalized. The year we chose for our comparison was 1976. The results are shown in Table 1. As can be seen, the rankings of the two data bases are reassuringly similar. The first and second in each ranking are the same. Indeed the first 12--which together account for over 95 percent of hospital discharge diagnoses in both samples--are the same, although the rank ordering differs slightly. The rank order correlation coefficient (0.90)

reflects this similarity. The proportions in the various diagnostic categories do vary somewhat--mostly notably the substantially higher incidence of pregnancy/childbirth among the AFDC sample, and the higher incidence of circulatory system problems among the National Health Survey females. This is to be expected, given the younger average age of the AFDC female sample.[1]

DESCRIPTIVE STATISTICS ON THE CALIFORNIA MEDICAID SAMPLES

The California Medicaid study made two major contributions in terms of sample classification. The first is classifying by predominant source of (ambulatory) care (PSC), on the argument that the PSC is the typical point of entry into the health care delivery system. The second is classifying by hospital discharge diagnosis according to the degree of discretion involved, which can be expected to shed some light both on the issue of case-mix differences and on the issue of possible hospitalizations from the outpatient department for economic (fill-the-bed) reasons in excess of what good medical practice would indicate. This section discusses these two classification schemes briefly and provides distributional data on our samples of AFDC females and the disabled in terms of these classifications.

The PSC classification scheme is a five-component mutually exclusive set of possibilities:

- Private Physician's Office
- Emergency Room
- Outpatient Department
- Clinic
- Nonuser of Ambulatory Care

---

[1]The higher incidence of pregnancy/childbirth for the AFDC female sample in this period may also be due in part to the fact that in California all females in single-headed households become eligible for Medicaid automatically when they become pregnant.

Table 1

Comparison of Diagnostic Data of a California Female AFDC Medicaid Sample and a Nationwide Female Sample 1976

(Percent of Total Cases)

Rank	Disease Grouping: National Health Survey[a] (females only)	Percent	Rank	California AFDC[b] (females only)	Percent
1.	Pregnancy/childbirth	19.78	1.	Pregnancy/childbirth	37.87
2.	Genitourinary System	11.69	2.	Genitourinary System	14.78
3.	Circulatory System	11.21	3.	Respiratory System	13.72
4.	Digestive System	10.75	4.	Digestive System	8.36
5.	Respiratory System	8.48	5.	Accidents	6.48
6.	Accidents	7.88	6.	Mental Disease	5.76
7.	Neoplasms	7.49	7.	Neoplasms	3.12
8.	Musculoskeletal	5.03	8.	Circulatory System	2.30
9.	Nervous System	4.00	9.	Infective parasitic disease	1.60
10	Mental Diseases	3.77	10.	Nervous System	1.25
11.	Endocrene/Nutrition	2.80	11.	Muskuloskeletal	1.24
12.	Infective parasitic diseases	2.26	12.	Endocrine System	1.15
13.	Symptoms	1.70	13.	Congenital anomalies	0.82
14.	Skin Diseases	1.44	14.	Symptoms	0.69
15.	Diseases of blood	0.85	15.	Skin Diseases	0.68
16.	Congenital anomalies	0.81	16.	Perinatal morbidity	0.16
17.	Perinatal morbidity	0.04	17.	Other	0.02
18.	Other	0.02	18.	Diseases of blood	0.00
	Total	100.00			100.00

[a]Based on discharge diagnoses from nonfederal short stay hospitals from the National Health Survey. Utilization of Short Stay Hospitals: Annual Summary for the United States, 1976, NCHS, USDHEW, Pub. No. (PHS) 78-1788.

[b]Based on a sample of 8,622 female AFDC beneficiaries of which 11% were hospitalized. Beneficiaries were not eligible for prepaid care but had to be Medicaid-eligible for 6 continuous months. Los Angeles County and drug claims are excluded. Those eligibles aged 7-14 or 65 or older are excluded. Data are weighted to correct for age cohort and geographic stratification.

For the private physician's office a further division was made, into General Practitioner or Specialist. For the hospital settings a further division was made, into public or private hospital. The PSC was defined as the setting where the maximum number of ambulatory physician visits occurred for a given patient during a given six-month period. Table 2 shows, for our two samples (the AFDC sample distribution is shown with and without pregnancies), hospital use and cost data. Table 3 shows, for our AFDC female sample, the distribution of the most frequent hospital diagnosis by PSC.

The physician discretion classification scheme was a six-component mutually exclusive set of possibilities. The diagnoses were then allocated to one of the categories according to expert medical judgment. Table 4 lists the six categories on the left, with representative examples of the diagnoses on the right. Table 5 shows the distributions of our AFDC female (with and without pregnancies) and disabled sample members by diagnosis group. Tables 6 and 7, respectively, show the distributions of AFDC (including pregnancies) and disabled sample members broken down by diagnosis group and PSC.

CONCLUSION

Medicaid claims data provide information on large samples at much lower cost than do other information sources of comparable detail. The purpose of this paper has been to give some idea of whether our analysis file gives an unbiased picture of the usage patterns of the Medicaid population, and whether that population is representative of other wider populations

With respect to the first, there are two major questions: Do first hositalizations reflect the distribution of all hospitalizations during a six-month period? Does the methodology used to assign a predominant source of ambulatory care produce an accurate measure of the incidence of visits to a private physician's office visits versus an outpatient department.

Reliability tests on both scores indicate that the calculated distributions are indeed unbiased. With respect to representativeness, comparison with diagnosis data from the National Health Survey shows very similar rankings by frequency of diagnosis.

Table 2

Hospitalization Data for a Sample of Medicaid Eligibles
by Eligibility Status, Type of Service, and Predominant Source of Care
California, 1976-1978

Eligibility Status	Type of Service	Predominant Source of Care[a]								
		Private Office		Public Hospital		Private Hospital			Nonuser of Ambulatory Care	Total
		GP	Specialist	ER	OPD	ER	OPD	Clinic		
AFDC[b] Female	Proportion hospitalized	0.10	0.14	0.24	0.20	0.13	0.18	0.07	0.04	0.11
	Average length of stay	4.61	5.17	2.07	5.16	5.59	3.52	6.34	3.46	4.80
	Mean expenditurea for inpatient care for those hospitalized	$1,257.32	$1,504.75	$500.69	$1,583.37	$1,624.64	$1,302.77	$2,226.44	$947.22	$1,393.59
	Mean expenditures/day for short-term inpatient care for those hospitalized	$269.23	$290.51	$242.23	$306.58	$290.61	$369.90	$351.22	$274.22	$290.50
	Sample Size	(2,174)	(2,997)	(102)	(467)	(491)	(234)	(253)	(1,904)	(8,622)
AFDC Females (without[c] delivery)	Proportion hospitalized	0.08	0.10	0.22	0.12	0.11	0.08	0.04	0.01	0.08
	Average length of stay	4.94	5.65	1.68	4.51	5.84	3.84	3.85	7.51	4.83
	Mean expenditures for short-term inpatient care for those hospitalized	$1,340.22	$1,632.34	$393.99	$1,109.39	$1,720.55	$1,256.21	$2,666.36	$1,202.81	$1,459.34
	Mean expenditures/day for ahort-term inpatient care for those hospitalized	$271.29	$289.13	$234.88	$246.24	$294.52	$326.56	$355.17	$249.19	$282.67
	Sample Size	(2,151)	(2,932)	(101)	(444)	(479)	(2230	(244)	(1,869)	(8,443)
Disabled[d]	Proportion hospitalized	0.11	0.20	0.29	0.19	0.31	0.22	0.12	0.05	0.13
	Average length of stay	8.04	10.60	13.00	9.78	7.69	14.95	10.85	13.25	10.35
	Mean expenditures for short-term inpatient care for those hospitalized	$1,480.21	$1,729.54	$2,124.49	$1,698.60	$986.70	$2,993.83	$2,385.16	$1,951.77	$1,726.77
	Mean expenditures/day for short-term inpatient care for those hospitalized	$184.22	$163.10	$163.42	$173.62	$128.38	$200.22	$219.83	$147.33	$166.89
	Sample Size	(1,939)	(2,354)	(133)	(433)	(263)	(306)	(144)	(2,578)	(8,151)

[a]Predominant source of care is defined as the setting where the maximum number of physican ambulatory visits occurred. OPD is outpatient department. ER is emergency room. GP is general practitioner.

[b]Based on a sample of 8,622 female AFDC Medicaid eligibles not eligible for prepaid care and eligible for 6 continuous months. Those aged 7-14 or 65 or older are excluded. Los Angeles county and drug claims are excluded. Data are weighted to correct for age cohort and geographic stratification. Those eligibles receiving long-term care are excluded from these data.

[c]Excludes women who had a delivery in the six-month period of January-June of the study year.

[d]Based on a sample of 8,151 male and female diaabled Medicaid eligibles. Thoae aged 65 or older are excluded. Los Angeles and drug claims are excluded. Those beneficiaries receiving long-term care are included in these data.

Table 3

Distribution of Fifteen Most Frequent Hospital Diagnoses by Predominant Source of Care for Female AFDC Medicaid Eligibles California, 1976-1978

Physician's Office		Outpatient Department	
Diagnosis[a]	Percent	Diagnosis[a]	Percent
Normal Delivery	13.4	Normal Delivery	17.9
Pneumonia	4.0	Pneumonia	4.7
Sterilization/healthy donor	3.8	Delivery with other complications	4.7
Acute bronchitis	3.3	Hypertrophy of tonsils/adenoids	2.5
Disorders of menstruation	3.1	Diarrheal disease	2.2
Diarrheal disease	2.2	Cholelithiasis	2.2
Hypertrophy of tonsils/adenoids	2.0	Cholecystitus	2.2
Delivery with other complications	1.9	Other complication of pregnancy	1.8
Disease of parametrium/pelvic peritoneum	1.9	Sterilization/healthy donor	1.4
		Bronchitis unqualified	1.4
Other disease of female genital	1.7	Other disease of urinary tract	1.4
Bronchitis unqualified	1.5	Neurosis	1.1[b]
Uterovaginal prolapse	1.4	Benign hypertension	1.1
Cholecystitus	1.2	Vertebrogenic pain syndrome	1.1
Diabetes Mellitus	1.1[b]	Symptoms referrable to abdomen/ lower GI	1.1
Other complication of pregnancy	1.1		
Cumulative Percent	43.6	Cumulative Percent	45.4

Emergency Room		Clinic	
Diagnosis[a]	Percent	Diagnosis[a]	Percent
Normal delivery	18.0	Normal delivery	41.9
Pneumonia	13.0	Single born - classification of infant	9.7
Other complication of pregnancy	6.2		
Bronchitis unqualified	3.1	Delivery with other complications	6.5
Delivery with other complications	3.1	Diarrheal disease	3.2[b]
Concussion	2.5	Infectious hepatitus	3.2
Adverse effect of other drugs	2.5	Gonoccal infections	3.2
Sterilization/healthy donor	2.5	Pediculus	3.2
Cholecystitus	1.9	Benign neoplasm of bone/cartilage	3.2
Symptoms referrable to abdomen/ lower GI	1.9	Neuroses	3.2
		Phlebitis/thrombophlebitis	3.2
Acute bronchitis	1.9	Other disease of cervix	3.2
Other viral diseases	1.2[b]	Uterovaginal prolapse	3.2
Benign neoplasm of ovary	1.2	Disorders of menstruation	3.2
Neuroses	1.2	Abortion for medical reasons	3.2
Sprains of the back	1.2	Dislocation of knee	3.2
		Burn	3.2
Cumulative Percent	60.2	Cumulative Percent	100.0

Note: Based on over 2,000 hospitalizations from a sample of approximately 22,000 female AFDC Medicaid eligibles not eligible for other prepaid care and eligible for 6 continuous months. Those 7-14 or 65 or older are excluded. Los Angeles and drug claims are excluded. Data do not correct for age cohort and geographic stratification. Predominant source of care is defined as the setting where the maximum number of physician ambulatory visits occurred for a given patient.

[a]Diagnostic codes are those of the International Classification of Diseases, adapted, Edition 8.

[b]Once the frequencies reach a certain level their distribution becomes flat with frequencies of only 1 or 2 per diagnosis. The last entries in the column, therefore, are examples of the many diagnoses with the same (small) frequency.

Table 4

Categorization of Physician Discretion
Based on Hospital Discharge Diagnosis

Category	Examples[a]
A. No discretion of whether to hospitalize and when, but also no indication of general health status	Normal delivery Most malignant neoplasms Meningitis
B. No discretion of whether to hospitalize or when, but correlated with poor health status	Cerebral hemorrhage Acute myocardial infarction Aortic aneurysm
C. Delayable hospitalization with quality of life choice/discretion of physician required	Hernia Cataract Hypertrophy of tonsils
D. Cannot judge based on data available on medical claims	Other bacterial diseases Other disease of thyroid gland
E. Usually involves discretion of physician as to how to treat	Diabetes mellitus Drug dependencies Acute bronchitis
F. Discretion of physician required; however, normally not hospitalized	Mumps Chickenpox Acute tonsillitis

[a]Disease descriptions from diagnosis codes in ICA8.

[b]A complete list of the ICDA8 codes by hospitalization category is available through Philip Held at The Urban Institute.

Table 5

Distribution of Hospital Discharge Diagnosis Groupings for California AFDC and Disabled Medicaid Eligibles 1976-1978

(percent)

Diagnosis Group	Eligibility Status		
	AFDC Females[a]		Disabled[c]
	All	Without Pregnancy[b]	
A. No discretion but also no indication of health status	39.10	15.60	20.51
B. No discretion but also correlated with poor health status	0.29	0.42	2.86
C. Delayable - Quality of life choice	10.04	14.63	7.31
D. Don't know; not enough information	20.22	25.23	27.83
E. Physician discretion about whether and how to treat	24.16	35.08	30.73
F. Usually not hospitalized	6.20	9.04	10.66
Total[d]	100.01	100.00	99.90

Note: Los Angeles and drug claims are excluded.

[a]Data based on a sample of 8,622 female AFDC eligibles not eligible for prepaid care and eligible for 6 continuous months. Those aged 7-14 years and those aged 65 years or older are excluded. Data are weighted to correct for age cohort and geographic stratification.

[b]Subgroup of the sample which excludes women who had a delivery in the six-month period of January through June in the study year.

[c]Data based on a sample of 8,151 male and female disabled eligibles; those aged 65 or older are excluded.

[d]May not add to 100 because of rounding.

Table 6

Distribution of Hospital Diagnoses Groupings by Predominant Source of Care, California Female AFDC Medicaid Recipients California, 1976-1978[a]

Percent

Diagnosis Group	Predominant Source of Care[c]							
	Private Office		Public		Private			
	GP	Specialist	ER	OPD	ER	OPD	Clinic	Total[d]
A. No discretion but also no indication of health status	28.84 (18.08)[b]	38.03 (17.10)	5.11 (5.61)	47.82 (5.81)	24.55 (13.91)	70.03 (30.16)	46.77 (4.15)	39.10 (15.60)
B. No discretion but also correlated with poor health	0.00 (0.00)	0.36 (0.51)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	7.25 (13.05)	0.29 (0.42)
C. Delayable - quality of life choice	8.90 (10.39)	13.33 (19.14)	0.00 (0.00)	3.44 (6.72)	13.63 (16.70)	4.45 (10.63)	0.00 (0.00)	10.04 (14.63)
D. Don't know; not enough information	18.18 (21.22)	21.09 (24.21)	83.58 (81.97)	15.85 (25.05)	21.57 (20.07)	7.71 (16.64)	9.18 (16.56)	20.22 (25.23)
E. Physician discretion about whether and how to treat	32.95 (38.47)	20.65 (29.65)	11.31 (12.42)	28.02 (52.90)	34.42 (42.18)	17.02 (40.68)	36.81 66.28	24.16 (35.08)
F. Usually not hospitalized	10.14 (11.83)	6.54 (9.39)	0.00 (0.00)	4.87 (9.52)	5.83 (7.14)	0.79 (1.89)	0.00 (0.00)	6.20 (9.04)
Total	100	100	100	100	100	100	100	100
Sample size	190 (167)	369 (304)	10 (9)	68 (57)	54 (42)	33 (22)	19 (10)	805 (626)

[a]Based on a sample of 8,622 female AFDC eligibles not eligible for prepaid care and eligible for 6 continuous months. Los Angeles county and drug claims are excluded. Those aged 7-14 or 65 or older are excluded. Data are weighted to correct for age cohort and geographic stratification. Eligibles who received care in a long-term care setting are not included in these data.

[b]Numbers in parentheses are for a subsample of (a) which excludes women who had a delivery during the six month period of January through June.

[c]Predominant source of care is defined as the setting where the maximum number of ambulatory physician visits occurred. GP is general practitioner. ER is emergency room. OPD is outpatient department.

[d]Nonusers of ambulatory care are included in the total.

Table 7

Distribution of Hospital Diagnoses Groupings by
Predominant Source of Care, California Disabled Medicaid Recipients
California, 1976-1978[a]

(Percent)

Diagnosis Group	Predominant Source of Care[b]							
	Private Office		Public		Private			
	GP	Specialist	ER	OPD	ER	OPD	Clinic	Total[c]
A. No discretion but also no indication of health status	15.89	15.89	21.05	20.88	35.40	24.66	21.05	20.51
B. No discretion but also correlated with poor health	3.49	2.87	0.00	1.10	0.00	1.37	5.26	2.36
C. Delayable - quality of life choice	7.75	8.17	2.63	4.40	1.61	4.11	15.79	6.81
D. Don't know; not enough information	27.13	29.58	28.95	26.37	17.74	27.40	10.53	27.13
E. Physician discretion about whether and how to treat	32.56	35.54	42.11	32.97	32.26	24.66	36.84	33.03
F. Usually not hospitalized	13.18	7.95	5.26	14.29	12.90	17.81	10.53	10.16
Total	100	100	100	100	100	100	100	100
Sample Size	258	453	38	91	62	73	9	1,102

[a]Based on a sample of 8,151 disabled eligibles. Los Angeles and drug claims are excluded. Those aged 65 or older are excluded. Eligibles who received care in a long-term care setting are included in these data.

[b]Predominant source of care is defined as the setting where the maximum number of ambulatory physician visits occurred. GP is general practitioner. ER is emergency room. OPD is outpatient department.

[c]Nonusers of ambulatory care are included.

CMS LIBRARY
3 8095 00012790 8

THE URBAN INSTITUTE
2100 M Street, N.W., Washington, D.C. 20037